Article

Hub-Periphery Hierarchy in Bus Transportation Networks: Gini Coefficients and the Seoul Bus System

Chansoo Kim ^{1,†}, Segun Goh ^{2,†} , Myeong Seon Choi ³, Keumsook Lee ^{4,*} and M. Y. Choi ^{5,*}

¹ AI Lab., Computational Science Center & ESRI, Korea Institute of Science and Technology, Seoul 02792, Korea; eau@econ.kist.re.kr

² Institut für Theoretische Physik II, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany; segun.goh@gmail.com

³ Graduate School of Nanoscience and Technology, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea; joym514@kaist.ac.kr

⁴ Department of Geography, Sungshin Women's University, Seoul 02844, Korea

⁵ Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul 08826, Korea

* Correspondence: kslee@sungshin.ac.kr (K.L.); mychoi@snu.ac.kr (M.Y.C.)

† Equal contributions.

Received: 1 August, 2020; Accepted: 2 September 2020; Published: 6 September 2020

Abstract: Bus transportation networks are characteristically different from other mass transportation systems such as airline or subway networks, and thus the usual approach may not work properly. In this paper, to analyze the bus transportation network, we employ the Gini coefficient, which measures the disparity of weights of bus stops. Applied to the Seoul bus system specifically, the Gini coefficient allows us to classify nodes in the bus network into two distinct types: hub and peripheral nodes. We elucidate the structural properties of the two types in the years 2011 and 2013, and probe the evolution of each type over the two years. It is revealed that the hub type evolves according to the controlled growth process while the peripheral one, displaying a number of new constructions as well as sudden closings of bus stops, is not described by growth dynamics. The Gini coefficient thus provides a key mathematical criterion of decomposing the transportation network into a growing one and the other. It would also help policymakers to deal with the complexity of urban mobility and make more sustainable city planning.

Keywords: bus transportation system; urban infrastructure; network growth; Gini coefficient; complex systems

1. Introduction

Complex networks have attracted much attention in various research areas. Since the seminal works on small-world networks [1] and scale-free networks [2], interest in spatial networks has grown significantly including various fields of application [3,4]. Until recent years, there has been much effort to assess the sustainability of transportation networks [5–11]. In view of sustainability, a better understanding of the network structure is required for improving those analyses [12–15]. Specifically, there have appeared studies of mass transportation networks: For instance, the network structure and passenger flows of the Seoul subway system were analyzed [16,17]. Further, the growth process of the Seoul subway network was described by a master equation for a Yule-type model [18], and the passenger flow on the network was described by a gravity model modified by a Hill function [19]. Another study includes the clique growth model, in which the growth of the network is governed

by attaching new cliques to the system [20]. In the case of the bus transportation network (BTN), scaling and renormalization ideas manifested the emergence of criticality in the passenger flow [21] while accessibility measurement was also successfully performed [22].

Among those spatial transportation networks explored, the BTN plays a substantial role in short-to mid-distance trips of passengers in a city. Therefore, changes in the BTN are rather important to network users (passengers) and network designers (policymakers). The BTN possesses some unique characteristics, different from those of other transportation systems [7,9,11,15,23]. Most of all, it serves periphery of a city with convenient and tangible transportation means to and from less noticed places in which other mass transportation systems such as the subway do not provide stops or stations [15]. Especially, in Seoul, while subway transportation has a major share for downtown areas including central business districts, residents in the suburbs take bus transportation for commuting. Part of the reason is that there are many mountainous and hilly areas in Seoul, which hinders the construction of subway stations at low cost [13]. In particular, the BTN changes, namely, vary in nodes (stops) and links (routes), usually much faster than others [24]. Furthermore, while a subway network has no noticeable hubs, the BTN has hubs and peripheral nodes that are distinct strikingly from each other.

Such characteristics make it rather inappropriate to apply existing models to the BTN [19–22]. To circumvent this difficulty, we in this paper propose a classification scheme to filter peripheral nodes out of the network with the aid of the disparity concept. The connectivity of each node, which is reflected in the weight, is compared; this manifests the duplex nature of the BTN and allows one to divide the nodes into two types, hubs and peripheries. We then probe the evolution of the two types from the viewpoint of the Yule process, which describes the well-known growth of an object growing proportionally to its current size. The Yule-type controlled growth process, extended and described generally by a master equation [25–27], was applied successfully to various growing complex systems [28,29]. It turns out that the growth process does not give a proper explanation for the evolution of the peripheral network displaying too many changes such as frequent constructions and closings of stops. Accordingly, it is indeed desirable to separate peripheral nodes from hubs, where the Gini coefficient [30,31] plays a key role. The Gini coefficient has long and widely been used in various fields [32,33]. There have been various interesting attempts to utilize it for separating data sets [34–36], recently including machine learning techniques [37,38].

The paper is organized as follows: Section 2 presents the Seoul bus transportation network, and we describe how one utilizes the Gini coefficient to assess the BTN. Each node in the BTN is categorized as a hub or a peripheral one, according to the Gini coefficient. The set of hub nodes composes the hub network, and in the same manner, the peripheral network is built of peripheral nodes. In Section 3, we probe how the hub network and the peripheral network changed over the two years, 2011 and 2013, and elucidate the evolution of the hub by means of the controlled growth mechanism. Section 4 summarizes our work and discusses implications for policymakers to overcome the vulnerability of the BTN as well as to improve it consistently.

2. Materials and Methods

As an exemplary BTN, we consider the Seoul bus system, which is a well-developed, large-scale BTN. We analyze the smart card (called *T-money card*) data, which are collected by the Seoul metropolitan government and not open to the public. The information contains the departure/arrival bus stops and times of each trip. Obtaining permission from the government, we have access to the dataset with the personal identification removed. Analyzing the passenger flow data for the years 2011 and 2013, we observe that the data as a whole resist explanation in terms of the controlled growth mechanism. We thus examine the hierarchical structure in the Seoul bus system by the help of the Gini coefficient, in advance of considering evolution as a controlled growth process.

2.1. Seoul Bus System

In the years 2011 and 2013, there were $N_s = 15,515$ and 15,702 bus stops, respectively, in the Seoul bus system, the spatial distributions of which are shown in Figure 1. Among those, 8408 stops turn out to be common in both years, as recognized by their identifications (IDs). On April 11th in the year 2011, 5,668,431 people in total rode on buses and only $N_e = 12,908$ stops out of the total have non-zero values of the strength. Here the strength s of a stop stands for the total number of passengers departing/arriving that stop, i.e., getting on/off buses at the stop, while the weight w_{ij} ($= w_{ji}$) between two stops i and j refers to the total number of passengers making trips between the two stops, i.e., getting on buses at stop i and off at stop j or vice versa. Accordingly, the weight w_{ij} and the strength s_i of stop i are related via: $s_i = \sum_j w_{ij}$.

Figure 1. Locations of bus stops in the Metropolitan Seoul bus system: Blue dots represent bus stops persisting between the two years 2011 and 2013. Bus stops present only in 2011 and in 2013 are represented by red and green dots, respectively. Numbers in parentheses refer to the number of corresponding stops.

We first probe the growth of the Seoul BTN with the aid of the Yule-type growth model, which was successfully applied to the Seoul subway network to reveal the emergence of log-normal, Weibull, and power-law distributions [18]. In sharp contrast, neither strength distributions nor weight distributions of the Seoul BTN appear consistent with the Yule-type growth model.

For instance, Figure 2 presents distributions of the strength data on 11 April 2011 and on 4 March 2013, together with the corresponding log-normal distributions predicted by the Yule-type growth model. As shown, the strength distributions exhibit conspicuous deviations, for small strengths ($s \lesssim 100$), from the log-normal and Weibull distributions as well as the power-law distribution. Assuming that such disagreement results from the hierarchy among the stops, we attempt to overcome this issue

by decomposing the bus stops into the two categories, hubs and peripheries, and considering the two separately. Namely, we presume that hubs and peripheries play different roles in the BTN. It is then plausible that the growth mechanism for these two types of nodes should be different from each other.

Figure 2. Distribution f of strengths s in the Seoul BTN, plotted on the semilogarithmic scale. Red squares and green circles are data points for the years 2011 and 2013, respectively. Solid and dotted lines plot the log-normal and the Weibull distributions, respectively, obtained via the maximum likelihood estimation from the data for $s \geq 25$. The location and the scale parameters of the log-normal distributions are given by $(\mu, \sigma) = (5.91, 1.27)$ for 2011 and $(5.92, 1.26)$ for 2013. The shape and the scale parameters of the Weibull distributions are estimated as $(\gamma, \eta) = (0.78, 700)$ for 2011 and $(0.79, 706)$ for 2013.

2.2. Gini Coefficients as a Classification Criterion

The strength simply quantifies the overall amount of passenger flows of a bus stop, and there is no *a priori* reason to prefer a criterion for categorization from a strength profile, for example, by setting an ad-hoc threshold value. Moreover, the location-based approach does not seem to be proper for the bus network, because peripheral nodes can also be located not only at the end of a route but also in the center of the city. Here we thus proceed one step further to characterize the flow distribution of each bus stop connected to other stops in the network. Specifically, we propose the *Gini coefficient*, which measures how evenly the passenger flows are distributed from one stop to others, to provide a sensible criterion serving the above purpose. The coefficient has been widely used to quantify inequality of a given system in view of distributions [30,31]. Suppose that in a society consisting of $N (\gg 1)$ people, the j th person possesses wealth x_j . The Gini coefficient for the society is then defined to be

$$G \equiv \frac{\sum_{j=1}^N \sum_{k=1}^N |x_j - x_k|}{2N \sum_{j=1}^N x_j}, \quad (1)$$

which ranges from zero (completely equal) to unity (completely unequal): $0 \leq G \leq 1$. It focuses on the average absolute difference between all pairs in the society.

We now consider, in place of x_j , the weight w_{ij} between given stop i and other stops j in the system. In principle, to reflect the connectivity structure of the BTN with meaningless pairs discarded, one needs to take into account bus routes serving the bus stop i . However, since buses always operate in one

direction at each bus stop and bus routes may have complicated topology in general [22], this turns out to be a demanding task. Therefore, in this study, we simply adopt the weight as an index assessing the connectivity between bus stops. This leads to the generalized Gini coefficient G_i , defined for every stop i in the BTN:

$$G_i \equiv \frac{\sum_{j=1}^{N_i} \sum_{k=1}^{N_i} |w_{ij} - w_{ik}|}{2N_i \sum_{j=1}^{N_i} w_{ij}}, \quad (2)$$

where N_i denotes the number of bus stops with non-zero passenger flows to and from the i th bus stop, i.e., bus stops j ($\neq i$) with $w_{ij} \neq 0$. The Gini coefficients computed for all bus stops in the Seoul BTN lead to the spatial distribution in Figure 3, where the Gini coefficient levels are color-coded. Overall, the spatial distributions in the years 2011 and 2013 turn out to be qualitatively the same, confirming that the Seoul BTN remained stable between these two years.

Figure 3. Heat map of Seoul depicting the spatial distribution of the Gini coefficient levels, specified by color, of all bus stops in the years (a) 2011 and (b) 2013. (For the color code, see the vertical bar on the right.) Accordingly, intense red regions correspond to those housing hub stops while the black line corresponds to the municipal boundary of Seoul.

Then the Gini coefficient G_i of each stop ($i = 0, 1, \dots, N$) is sorted in the decreasing order. In plotting data, we consider the Gini coefficient G at the increment of 0.01. Namely, we divide the G -axis into intervals of length 0.01 and take the average in each interval. The resulting distribution of the Gini coefficients for all stops is shown in Figure 4, where each data point represents the average over stops in each interval (of length 0.01) of the Gini coefficient.

Figure 4. Distributions of Gini coefficients in 2011 (red) and 2013 (green). Lines are merely guides to the eye.

3. Results

3.1. Hub and Peripheral Nodes

We make use of the Gini coefficient to decompose the BTN into two groups of nodes: hubs and peripheries. Usually, peripheral nodes are connected to only a few (two in the majority of the case) nodes, since many of such peripheral stops are served by single routes. In consequence, the weight w_{ij} from a peripheral stop i may not vary much with j except those linked directly via the route serving i . This should result in a relatively small value of the Gini coefficient G_i . On the other hand, hubs, each served by a relatively large number of routes, accommodate many passengers with the numbers depending on the destination stops. Namely, the weight w_{ij} from a hub stop i is expected to vary significantly with j according to whether stop j is linked directly via routes serving i . In short, hubs and peripheries are expected to be characterized by relatively large and small values of the Gini coefficient, respectively, located in the red regions and in the blue to yellow regions in Figure 3.

To confirm this, we compute the mean weight $\bar{w}$ of each node. The mean weight $\bar{w}_i$ of stop i is given by the average of weight w_{ij} over the origin/destination stop j : $\bar{w}_i \equiv d_i^{-1} \sum_{j=1}^{d_i} w_{ij} = s_i / d_i$ with the degree d_i being the number of origin and destination stops to and from stop i , respectively. The mean weight versus the Gini coefficient is presented in Figure 5. It is observed that the mean weight tends to increase with the Gini coefficient, which suggests that passengers using a stop tend to increase more than proportionally with the number of origins/destinations served by the stop. Since hub stops are in general expected to have larger degrees and weights (i.e., serve more destinations and passengers) in comparison with peripheral stops, it is sensible that hubs/peripheries have relatively large/small values of the Gini coefficient.

The remaining task is how to choose the boundary between hubs and peripheries. Here we notice that Figure 5 exhibits characteristic changes of the behavior at the Gini coefficient $G \approx 0.5$: Specifically, the mean weight growth with G changes its slope at $G \approx 0.5$. For more information, we also consider the rank distribution of the Gini coefficient and present the result in Figure 6, where each data point represents the Gini coefficient at each rank. To probe the change in the decreasing behavior of the Gini coefficient with the rank, we perform the convexity test [39], which reveals that the convexity changes (from convex to non-convex behavior) around the rank 11,000. It is pleasing that this rank corresponds to $G \approx 0.5$ as well. This apparently suggests $G \approx 0.5$ as the boundary between hubs and peripheries. As a result, we choose specifically the rank 11,108 as the criterion separating hubs and peripheries since the bus stop ID on this rank commonly exists in both 2011 and 2013 data sets.

Figure 5. Mean weight $\bar{w}$ versus the Gini coefficient G for each node. The two lines at each year, least-square fits of the data points below and above $G \approx 0.5$, meet at that point.

Figure 6. Rank distribution of the Gini coefficient. The convexity is shown to change around the rank 11,000.

3.2. Weight Changes Depending on the Node Type

To probe the time evolution of the Seoul BTN, we first consider the change of the weight between nodes (stops) of the two types during the two years from 2011 to 2013 and analyze the application of the Yule-type growth model to the data sets. Specifically, comparison of the data in 2011 (April 11th) and in 2013 (March 4th) shows increases both in bus stops (from 15,515 to 15,702) and in passengers (from 5,668,431 to 6,188,333). Comparing bus stop IDs in both years, we identify 8408 stops common in both years, as presented in Figure 1; these correspond to 64.1% of total stops in 2011 and 63.9% of those in 2013. In extracting common bus stops, we have considered bus stop IDs and GPS locations,

and adopted the bus stop ID as the criterion. It turns out that even for the same GPS location, the stop IDs vary frequently during the two years. Reasons for such variations are not released publicly and are presumed to be policy and urban planning issues.

We now consider weights between any two among the stops, classified into the three types of pair: periphery-periphery, hub-periphery, and hub-hub pairs. Then several regression methods, including the least square regression (LSR), least absolute deviation (LAD) [40], and orthogonal regression (OR) [41–43], are applied, leading to the correlations of the three types of weight between the two years. Figure 7 exhibits the obtained weight correlations between the years 2011 and 2013. The slopes of the correlations obtained via the three regression methods are presented in Table 1.

Table 1. Slopes of the weight changes obtained via regression analyses of the data for the two years, 2011 and 2013.

	LSR	LAD	OR
periphery-periphery	0.93	0.91	0.72
Hub-periphery	0.95	0.97	0.71
Hub-hub	0.94	0.97	0.72

Figure 7. Weight correlations between the years 2011 and 2013. Weights between (a) periphery-periphery pairs, (b) hub-periphery pairs, and (c) hub-hub pairs. Blue solid lines plot the results of least square regression (LSR). Note the difference in the scale of the box.

The weight correlations between peripheral nodes show a bit smaller values of the slope (albeit obscure in OR), compared with the other two cases involving hub nodes. This indicates that, albeit not very conspicuous, the weights between peripheral nodes have decreased more than those involving hub nodes. One may argue that such a downward tendency in the weight between peripheral nodes is not consistent with the growth model, which usually predicts the emergence of power-law or log-normal type behavior. It is a consequence of the fact that there were far more stops in 2013 than in 2011, which reflects the construction of new stops. Despite the growth of the total strength (from 5,668,431 to 6,188,333), the average strength (per stop) of the peripheries turns out to have decreased by 30.2% while that of the hubs increased by 17.4%.

3.3. Evolution of the Seoul Bus Network

The bus stop rank sorted by the Gini coefficient provides a criterion for discriminating stops in the BTN into two distinct types: hubs and peripheries, which allows us to analyze the distributions and growth mechanisms for the two separately. Specifically, those bus stops with ranks above 11,108 are classified as hubs, while lower-ranked bus stops, having ranks below 11,108, constitute peripheral nodes.

We first examine the weight distribution of each category, which supports the validity of the Gini coefficient-based categorization. Indeed Figure 8 shows that the weight distributions fit well to either log-normal or power-law distributions. In each case, the parameters are estimated best via ordinary least squares (OLS) with the root-mean-square (RMS) deviation. For the log-normal distribution, they also agree with those via the maximum likelihood estimation (MLE). The distributions of weights

involving peripheral nodes appear to follow power-law distributions: $f(x) \sim x^{-\beta}$, with the exponent $\beta = 2.17$ and 2.23 for the weights between two peripheries and between a hub and a periphery, respectively, in the year 2013. In 2011, we have $\beta = 2.18$ for the weights between two peripheries and 2.22 for those between a hub and a periphery. On the other hand, weights between hubs exhibit a log-normal distribution with the location and the scale parameters $(\mu, \sigma) = (1.74, 1.34)$ for 2011 and $(1.75, 1.37)$ for 2013.

These results indicate that while the BTN is constructed, passengers tend to make trips to or from newly constructed peripheral stops. The corresponding production of new weights gives rise to the power-law behavior of the resulting weight distributions, with the exponents β greater than unity, according to the growth model [19]. In contrast, as to hub stops, the growth of the existing trips between them is dominant, leading to the log-normal distribution of the hub-hub weights.

Equipped with the finding from the analysis of the weight distributions, we also probe the growth mechanism for the strength distributions, shown in Figure 9. It is observed that the strength distributions of the hub category follow log-normal distributions. Notably, the deviations in the small strength regime in Figure 2 are not present. The strength distributions of peripheral stops also fit log-normal distributions in the range $s \gtrsim 100$. Table 2 summarizes the distribution parameters for the two years 2011 and 2013.

Table 2. Location and scale parameters, μ and σ , for the log-normal distribution, with their confidence intervals (CI) in 2011 and 2013.

	Year	μ	CI(μ)	σ	CI(σ)
Hub	2011	5.95	5.90 ~ 5.99	1.51	1.48~1.55
	2013	6.34	6.29~6.39	1.52	1.48~1.55
Periphery	2011	5.86	5.70~5.75	1.46	1.49~1.52
	2013	5.58	5.82~5.89	1.53	1.44~1.49

Figure 8. Distributions of (a) periphery-periphery, (b) hub-periphery, and (c) hub-hub weights. In (a,b), the red lines represent power-law distributions with exponent $\beta = 2.17$ and 2.18 , respectively, for the year 2011; green lines plot power-law distributions for 2013, with exponent $\beta = 2.23$ and 2.22 , respectively. In (c) the red and green lines depict the log-normal distributions with $(\mu, \sigma) = (1.74, 1.34)$ for 2011 and $(1.75, 1.37)$ for 2013, respectively. In all cases, the two lines overlap with each other too closely to distinguish.

Figure 9. Strength distributions on the log-log scale. Hub nodes in (a) 2011 and (b) 2013; peripheral nodes in (c) 2011 and (d) 2013. Solid lines plot the log-normal distributions with parameters given in Table 2.

In the perspectives of the controlled growth model, the emergence of log-normal distributions may imply the presence of a certain evolution mechanism such as growth and production in the BTN. In the case of hub nodes, time evolution is governed by the growth of each node; on the other hand, production (birth) of new nodes should be taken into account to understand the time evolution of peripheral nodes.

To probe the time evolution of the Seoul BTN, we consider the bus stops having the same ID in Figure 1. It is known that the Yule-type growth (possibly with self-size production) results in the log-normal distribution [26,29]:

$$f(s, t) = \frac{1}{\sqrt{2\pi}\sigma_t s} \exp \left[-\frac{(\ln s - \mu_t - \mu_0)^2}{2\sigma_t^2} \right], \quad (3)$$

where the location parameter μ and the scale parameter σ evolve in time according to

$$\begin{aligned} \mu_t &= \lambda t \ln(1 + b) \\ \sigma_t &= \sqrt{\lambda t} \ln(1 + b) \end{aligned} \quad (4)$$

with the (mean) growth rate λ , growth factor b , and initial location μ_0 .

We now examine the Seoul BTN in view of Equation (3), to check the applicability of the Yule-type growth model [26]. Analyzing the strength data for the hubs in the two years 2011 and 2013, we find the growth rate $\lambda = 0.496$ and growth factor $b = 0.129$. Since we focus on bus stops common in both years, it is natural that the growth probability for two years, given by $\lambda\Delta t$ with $\Delta t = 2$ (years), takes the value close to unity. Plugging these values into Equation (4), we obtain the changes in the location parameter and the scale parameter during two years: $\Delta\mu = \lambda\Delta t \ln(1 + b) = 0.120$ and $\Delta\sigma^2 = \lambda\Delta t [\ln(1 + b)]^2 = 0.0146$. These theoretical results are to be compared with the results from

the passenger data, $\Delta\mu = 0.39$ and $\Delta\sigma^2 = 0.0097$. Unlike the scale parameter, the location parameter shows a substantial discrepancy between the theoretical result and passenger data. In consideration of the limited data with the quite small growth factor, however, the two results, one from theoretical analysis and the other from passenger data, are presumed reasonably consistent with each other.

On the other hand, the growth of peripheral nodes is not captured properly by the controlled growth model. In particular, the peripheries display rather shrinking behavior than growth, as the passenger data have the mean $\mu = 5.58$ in 2013, decreased from $\mu = 5.86$ in 2011. Such different behaviors between hubs and peripheries can be attributed to the difference in bus stops of common IDs over the two years. Among those common stops, hub nodes are fully consistent over the years 2011 and 2013. Accordingly, each hub is subject to growth, albeit without newly-constructed bus stops, and the set of hubs is described by the controlled growth model, giving rise to the log-normal distribution. In contrast, only about half of the peripheral nodes persisted and are included in the set of common stops. To be precise, 48.71% of the peripheries present in 2011 were closed during the two years from 2011 to 2013 while newly constructed nodes during the two years occupied 48.04% of the total peripheries in 2013. We also attempt to match those peripheral stops of changed IDs with the help of their GPS locations, which turns out to give about 50% of stops persistent as well; in the case of hubs, this percentage of unchanged stops reaches 65% of the total. This analysis manifests that bus stops having small values of the Gini coefficient change frequently due to the low usage.

This also provides, as a by-product, a good reason to name this set of nodes *peripheral*, which are vulnerable to frequent changes at the policy level. Planning policies could include various government-creating or artificial transitions to remove less-visited stops. As Seoul is a very mature metropolitan city, its downtown area is already sufficiently saturated [19]. In consequence, closing of less used stops or construction of new stops is naturally concentrated in peripheral parts of the city. It should also be noted that the peripheral parts are not necessarily confined within suburban areas; stops in the area with less-developed transportation are possibly categorized as peripheral nodes.

4. Discussion

The BTN has unique characteristics, which may not be easily captured as in other transportation systems such as subways or airlines. Serving as a dominant transportation mode for short- to mid-distance trips, the BTN is important for policymakers to deal with the complexity of urban mobility and make more sustainable city planning, which should impact on both passengers and policymakers. To apply sustainability analysis with better accuracy, one needs to understand the BTN concerning not only distribution but also growth. Unlike other transportation networks, the passenger flow in the BTN may not be described plausibly by a single distribution, which reflects that there are two distinct types of bus stops and the trips thus vary according to the type involved.

In this study, we have considered how to characterize the BTN with regard to its evolution in time; this is expected to provide a fundamental approach to a better understanding of its structure, which should help sustainability assessment. Possible candidates for characterization include locations of bus stops, the strength of each stop, weight between a pair of stops, passenger concentration, and so on. We have here proposed to use the Gini coefficient, which measures disparity among weights of trips between each pair of nodes. Making use of the Gini coefficient for each bus stop, we have classified the stops into two distinct types: hub and peripheral nodes. Separating hubs having higher ranks and peripheries with lower ranks, we have observed log-normal distributions for both sets of nodes.

Specifically, comparison of the data in the years 2011 and 2013 has shown strengths of the hubs to grow, which is well explained by the Yule-type growth model. Analyzing the passenger data in the Seoul BTN, we have obtained parameters for the model. It has thus been concluded that the controlled growth model well captures the evolution of the set of hub nodes: It provides a good description, not only qualitatively but also quantitatively, of such features as the log-normal distribution, evolving through the mean and the deviation in time.

On the other hand, peripheral nodes are vulnerable to frequent changes due to low usage, which makes it inappropriate to apply the growth model. In Seoul, where downtown areas are already substantially saturated, changes in the BTN occur more often in peripheral nodes. To summarize, we point out that the Gini-based approach captures the key characteristics of the BTN among other transportation networks. It would be desirable to develop an algorithm to match closed bus stops with newly-constructed ones. Here, appropriate criteria may include not only the GPS locations of stops but also some policy-level information. This is left for future study.

Author Contributions: Conceptualization, K.L. and M.Y.C.; data curation, C.K., S.G., M.S.C., and K.L.; formal analysis, C.K., S.G., and M.S.C.; funding acquisition, K.L. and M.Y.C.; investigation, K.L. and M.Y.C.; methodology, S.G.; project administration, M.Y.C.; resources, K.L.; software, C.K., S.G., and M.S.C.; supervision, M.Y.C.; validation, C.K. and K.L.; writing—original draft, S.G. and K.L.; writing—review and editing, C.K. and M.Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Sungshin Women's University Research Grant of 2018 (KL), by the National Research Foundation of Korea through the Basic Science Research Program (Grant No. 2019R1F1A1046285) (MYC), and by Korea Institute of Science and Technology through the Economic and Social Science Research Initiative (ESRI) 2E30443 and Pride-up 2V08460 (CK).

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Watts, D.; Strogatz, S. Collective of small-world networks. *Nature* **1998**, *393*, 440–442. [[CrossRef](#)]
2. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [[CrossRef](#)] [[PubMed](#)]
3. Acemoglu, D.; Carvalho, V.M.; Ozdaglar, A.; Tahbaz-Salehi, A. The network origins of aggregate fluctuations. *Econometrica* **2012**, *80*, 1977–2016. [[CrossRef](#)]
4. Jackson, M.O. *Social and Economic Networks*; Princeton University Press: Princeton, NJ, USA, 2008.
5. Pardee, F. United States. Northeast Corridor Transportation Project. In *Measurement and Evaluation of Transportation System Effectiveness*; Rand Corporation Memorandum RM-5869-DOT; Rand Corporation: Santa Monica, CA, USA, 1969.
6. Altshuler, A.; Womack, J.; Pucher, J. *The Urban Transportation System: Politics and Policy Innovation*; MIT Press Series in Transportation Studies; MIT Press: Cambridge, MA, USA, 1979.
7. Dubois, D.; Bel, G.; Llibre, M. A set of methods in transportation network synthesis and analysis. *J. Oper. Res. Soc.* **1979**, *30*, 797–808. [[CrossRef](#)]
8. Zegras, C. Sustainable transport indicators and assessment methodologies—Biannual Conference and Exhibit of the Clean Air Initiative for Latin American Cities. In *Proceedings of the Sustainable Transport: Linkages to Mitigate Climate Change and Improve Air Quality*, São Paulo, Brazil, 25–27 July 2006.
9. Hu, Y.; Zhu, D. Empirical analysis of the worldwide maritime transportation network. *Physica A* **2009**, *388*, 2061–2071. [[CrossRef](#)]
10. Canas, G.; Rosasco, L. Learning probability measures with respect to optimal transport metrics. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: New York, NY, USA, 2012; pp. 2492–2500.
11. Farahani, R.Z.; Miandoabchi, E.; Szeto, W.; Rashidi, H. A review of urban transportation network design problems. *Eur. J. Oper. Res.* **2013**, *229*, 281–302. [[CrossRef](#)]
12. Zietsman, J.; Rilett, L.; Kim, S. *Sustainable Transportation Performance Measures for Developing Communities*; Technical Report; Texas Transportation Institute, Texas A&M University System: San Angelo, TX, USA, 2003.
13. Kim, H. *Developing Guidelines for Evaluating and Selecting Urban Rail Transit Systems*; The Korea Transport Institute: Seoul, Korea, 2008.
14. Jeon, C.M.; Amekudzi, A.A.; Guensler, R.L. Sustainability assessment at the transportation planning level: Performance measures and indexes. *Transp. Policy* **2013**, *25*, 10–21. [[CrossRef](#)]
15. Liu, X.; Derudder, B.; Wu, K. Measuring polycentric urban development in China: An intercity transportation network perspective. *Reg. Stud.* **2016**, *50*, 1302–1315. [[CrossRef](#)]
16. Lee, K.; Jung, W.S.; Park, J.S.; Choi, M.Y. Statistical analysis of the Metropolitan Seoul subway system: Network structure and passenger flows. *Physica A* **2008**, *387*, 6231. [[CrossRef](#)]

17. Lee, K.; Park, J.S.; Choi, H.; Jung, W.S.; Choi, M.Y. Sleepless in Seoul: The ant and the metrohopper. *J. Korean Phys. Soc.* **2010**, *57*, 823.
18. Lee, K.; Goh, S.; Park, J.S.; Jung, W.S.; Choi, M.Y. Master equation approach to the intra-urban passenger flow and application to the metropolitan Seoul subway system. *J. Phys. A* **2011**, *44*, 115007. [[CrossRef](#)]
19. Goh, S.; Lee, K.; Park, J.S.; Choi, M.Y. Modification of the gravity model and application to the metropolitan Seoul subway system. *Phys. Rev. E* **2012**, *86*, 026102. [[CrossRef](#)] [[PubMed](#)]
20. Yang, X.H.; Chen, G. Bus transport network model with ideal n-depth clique network topology. *J. Phys. A* **2011**, *390*, 4660–4672.
21. Goh, S.; Lee, K.; Choi, M.Y.; Fortin, J.Y. Emergence of criticality in the transportation passenger flow: Scaling and renormalization in the Seoul bus system. *PLoS ONE* **2014**, *9*, e89980. [[CrossRef](#)] [[PubMed](#)]
22. Lee, K.; Park, J.S.; Goh, S.; Choi, M.Y. Accessibility measurement in transportation networks and application to the Seoul bus system. *Geograph. Anal.* **2019**, *51*, 339. [[CrossRef](#)]
23. Stenneth, L.; Wolfson, O.; Yu, P.S.; Xu, B. Transportation Mode Detection Using Mobile Phones and GIS Information. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 54–63.
24. Badia, H. Comparison of bus network structures in face of urban dispersion for a ring-radial city. *Netw. Spat. Econ.* **2020**, *20*, 233–271. [[CrossRef](#)]
25. Choi, M.Y.; Choi, H.; Fortin, J.Y.; Choi, J. How skew distributions emerge in evolving systems. *EPL* **2009**, *85*, 30006. [[CrossRef](#)]
26. Goh, S.; Kwon, H.W.; Choi, M.Y.; Fortin, J.Y. Emergence of skew distributions in controlled growth processes. *Phys. Rev. E* **2010**, *82*, 061115. [[CrossRef](#)]
27. Choi, J.; Choi, M.Y.; Yoon, B.G. Interaction effects on the size distribution in a growth model. *J. Korean Phys. Soc.* **2018**, *72*, 327. [[CrossRef](#)]
28. Kwon, H.W.; Choi, M.Y. How cells grow and divide: Mathematical analysis confirms demand for the cell cycle. *J. Phys. A* **2012**, *45*, 135101. [[CrossRef](#)]
29. Kim, C.; Kim, D.S.; Ahn, K.; Choi, M.Y. Dynamics of analyst forecasts and emergence of complexity: Role of information disparity. *PLoS ONE* **2017**, *12*, e0177071. [[CrossRef](#)] [[PubMed](#)]
30. Gini, C. *Variabilità e Mutabilità*; Libreria Eredi Virgilio Veschi: Rome, Italy, 1912.
31. Gini, C. Measurement of inequality of incomes. *Econ. J.* **1921**, *31*, 124–126. [[CrossRef](#)]
32. Lambert, P.J.; Aronson, J.R. Inequality decomposition analysis and the Gini coefficient revisited. *Econ. J.* **1993**, *103*, 1221–1227. [[CrossRef](#)]
33. Dorfman, R. A formula for the Gini coefficient. *Rev. Econ. Stat.* **1979**, *61*, 146–149. [[CrossRef](#)]
34. Giger, E.; Pinzger, M.; Gall, H. Using the Gini coefficient for bug prediction in eclipse. In Proceedings of the 12th International Workshop on Principles of Software Evolution and the 7th Annual ERCIM Workshop on Software Evolution, Szeged, Hungary, 5–6 September 2011; IWPSE-EVOL '11, ACM: New York, NY, USA, 2011; pp. 51–55. [[CrossRef](#)]
35. Mussard, S.; Terraza, M.; Seyte, F. Decomposition of Gini and the generalized entropy inequality measures. *Econ. Bull.* **2003**, *4*, 1–6.
36. Lisker, T. Is the Gini coefficient a stable measure of galaxy structure? *Astrophys. J. Suppl. Ser.* **2008**, *179*, 319–325. [[CrossRef](#)]
37. Sundhari, S. A knowledge discovery using decision tree by Gini coefficient. In Proceedings of the 2011 International Conference on Business, Engineering and Industrial Applications, Kuala Lumpur, Malaysia, 5–7 June 2011; pp. 232–235.
38. Sanasam, R.; Murthy, H.; Gonsalves, T. Feature selection for text classification based on Gini coefficient of inequality. In Proceedings of the Fourth International Workshop on Feature Selection in Data Mining, Hyderabad, India, 21 June 2010; Liu, H., Motoda, H., Setiono, R., Zhao, Z., Eds.; PMLR: Hyderabad, India, 2010; Volume 10, Proceedings of Machine Learning Research, pp. 76–85.
39. Boyd, S.; Vandenberg, L. *Convex Optimization*; Cambridge University Press: New York, NY, USA, 2004; pp. 324–342.
40. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **1991**, *7*, 186–199. [[CrossRef](#)]

41. Carroll, R.J.; Ruppert, D. The use and misuse of orthogonal regression in linear errors-in-variables models. *Am. Stat.* **1996**, *50*, 1–6.
42. Maronna, R. Principal components and orthogonal regression based on robust scales. *Technometrics* **2005**, *47*, 264–273. [[CrossRef](#)]
43. Leng, L.; Zhang, T.; Kleinman, L.; Zhu, W. Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science. *J. Phys. Conf. Ser.* **2007**, *78*, 012084. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).